This is a paper of basic interest and shows that technical matters can be presented with intelligent humor. The author points out that a statistical binomial model used to determine the consistency of observations with a genetic hypothesis has a low discriminatory ability. The reasons for this performance are discussed.

A METHODOLOGICAL PROBLEM IN TESTING A RECESSIVE GENETIC HYPOTHESIS IN HUMAN DISEASE

Abraham M. Lilienfeld, M.D., M.P.H., F.A.P.H.A.

THE DETERMINATION of whether a given human disease is inherited is often quite difficult. In genetic studies the three available approaches are:

1. The study of the familial aggregation of the disease
2. Twin studies
3. The study of the frequency of consanguineous matings among parents of individuals with the disease.

Here I shall discuss one aspect of the study of familial aggregation in which we have been interested during the past few years. The other approaches are mentioned to indicate that the present discussion is not all-inclusive and is limited to one specific method of determining genetic factors in human disease.

We realize that the determination of familial aggregation of a human disease is not necessarily indicative of genetic factors, since it may be a result of environmental factors common to family members. However, the presence of familial aggregation is obviously consistent with a genetic hypothesis. To distinguish between genetic and environmental causes, investigators try to eliminate possible environmental etiological factors and to increase the specificity of the consistency of the observations with a genetic hypothesis. Thus, attempts are made to determine whether the observations agree with a dominant-gene hypothesis, a recessive-gene hypothesis, sex-linked genes, and so forth. This has led to the development of various statistical technics to test these specific hypotheses, the technics depending on the type of data collected. One such test was used by Steinberg and Wilder[1] in analyzing the inheritance of diabetes mellitus, which was similar to one used by Allan in 1933[2] to study the same disease entity, although it has not been used to any great extent during the intervening years.

Steinberg and Wilder presented data collected from a clinic group of diabetic patients who supplied information on the age and diabetes history of their parents and siblings. The data were used to test the hypothesis that diabetes is due to a single recessive gene under the following assumptions: (1) random mating, (2) equal fertility of all types of matings, (3) equal viability of all offspring, and (4) independence of age of

Table 1—Derivation of binomial model for testing recessive gene hypothesis

No. of Affected Parents	Parental Genotypes	Population Frequency of Parental Matings by Genotypes (a)	Frequency of Parental Matings in Total Population Who Are Capable of Producing Affected Offspring (b)	Proportion of Recessive Offspring in Each Mating Class (c)	Frequency in Total Population of Mating Class of Parents of Affected Offspring (d) = (c) x (b)	Proportion of Each Mating Class Among Parents of Affected Offspring (e) = (d)/p^2
0	AAxAA	q^4				
	AAxAa	$4q^3p$				
	AaxAa	$4p^2q^2$	$4p^2q^2$	1/4	p^2q^2	q^2
1	AAxaa	$2p^2q^2$				
	Aaxaa	$4p^3q$	$4p^3q$	1/2	$2p^3q$	$2pq$
2	aaxaa	p^4	p^4	1	p^4	p^2

onset of diabetes in the affected offspring from type of parental mating.

Assuming that the clinic population is equivalent to a random sample of diabetic offspring from a general population. the following test of the simple recessive-gene hypothesis is made:

1. The number of each of the three types of parental matings is determined, i.e.. the number of patients, neither of whose parents had diabetes; the number of patients, one of whose parents had diabetes; and the number, both of whose parents had diabetes. These represent the observations.

2. The gene frequency, denoted by "p." is determined from the frequency of diabetes among the parents of the diabetic patients by standard methods.

3. The expected proportion of each of the mating types, i.e., expected on the basis of the recessive-gene hypothesis, should be q^2, 2pq, and p^2 for the neither parent affected, one parent affected, both parents affected matings, respectively, where $q = 1-p$.

4. The observed distribution of parental mating types is tested by means of a chi-square test to see if it agrees with the expected distribution.

The derivation of the test—which we shall call the "binomial test"—is presented in Table 1. It is based on the Hardy - Weinberg equilibrium, from which the frequencies of those parental matings capable of producing affected offspring are determined. These frequencies are then reduced by the ratios 1/4, 1/2, and 1 to obtain the frequency in the population of the types of parental mating classes among affected offspring. The proportion of each parental mating class among the parents of the affected offspring is then determined by division of the previously obtained frequencies by p^2. The resultant frequencies are binomially distributed. Essentially, this model determines the expected proportion of the three different possible types of parental mating among a group of affected individuals, providing the disease was determined by a recessive gene.

Steinberg and Wilder applied this test to other published data on diabetes. In Table 2 we have reproduced from their paper a table containing the results of this application. Note that of four sets of data three fit the recessive gene model, as indicated by P values of over 0.05, whereas one does not, the P value being less than 0.01.

We were interested in the adequacy of this particular test in discriminating between genetic and nongenetic factors,

since we thought that it might be possible to obtain a similar kind of model on nongenetic assumptions. Paul Meir and I looked at this possibility on theoretical grounds and it became apparent that some nongenetic causes of familial aggregation may be represented by a binomial model. However, the most reasonable way to study this was to obtain similar data on an individual characteristic which occurred with a certain degree of familial aggregation primarily due to nongenetic, social, or environmental conditions. These data would be analyzed as if one were attempting to determine whether this characteristic was inherited as a recessive gene by means of the binomial test. Two rather similar characteristics were investigated in this manner.

The first characteristic we were able to study was attendance at medical school. By means of a questionnaire all students attending the University of Buffalo Medical School were asked whether or not their parents were physicians. The questionnaire was completed by 261 students; about 15 to 20 students did not complete or hand in the questionnaire. Thus, the characteristic was considered as attending or having attended a medical school sometime during an individual's life span. Table 3 contains the distribution of the students by the fol-

Table 2—Comparison of expected frequencies of matings yielding diabetic offspring with observed frequencies for four sets of data*

Mating	Steinberg and Wilder		Pincus and White		Allan		Harris	
	Expected	Observed	Expected	Observed	Expected	Observed	Expected	Observed
Both parents diabetic	21.6	22	3.6	3	0.8	2	3.1	8
One parent diabetic	370.8	370	78.8	80	19.4	17	118.8	109
Neither parent diabetic	1,588.6	1,589	440.6	440	122.8	124	1,119.1	1,124
X^2 (1 d. f.)	0.009		0.119		2.109		8.573	
P	>0.90		>0.70		>0.10		<0.01	

* From Steinberg and Wilder[1], p. 127.

Table 3—Comparison of observed and expected numbers of parental mating classes among offspring attending medical school

Parental Mating Class	No. of Medical Students in Mating Class	
	Observed	Expected
Both parents physicians	0	1.0
One parent physician	32	30.0
Neither parent physician	229	230.0
Total	261	261.0
p = "gene frequency"	0.0613	
Chi-square (1 d.f.)	1.14	
P	0.30>P>.20	

Table 4—Comparison of observed and expected numbers of parental mating classes among offspring attending the University of Buffalo

Parental Mating Class	No. of Students in Each Mating Class	
	Observed	Expected
Both parents attended University of Buffalo	14	4.4
One parent attended University of Buffalo	232	251.6
Neither parent attended University of Buffalo	3,640	3,630.0
Total	3,886	3,886.0
p = "gene frequency"	0.0335	
Chi-square (1 d.f.)	22.51	
P	<0.001	

lowing three parental classes: (1) neither parent had attended a medical school, (2) one parent had attended a medical school, and (3) both parents had attended a medical school. From these data the equivalent to the gene frequency, p, was estimated as 0.0613. Expected numbers based on a binomial distribution were obtained and a chi-square test of the goodness of fit of the observed numbers with expectations resulted in a probability level of between 0.2 and 0.3. From a purely formalistic viewpoint, one could then make the inference from these data that attendance at medical school was consistent with a recessive gene hypothesis. It is quite apparent that the test of the genetic hypothesis was not able, in this instance, to distinguish a type of familial aggregation that is principally determined by sociocultural factors.

Since the number of individuals in the sample was small and the parental mating class with two affected parents was absent, we selected another characteristic for a similar analysis. This characteristic was defined as attendance at the University of Buffalo and all students attending the university are considered as being affected. At the time of registration at the university all students are asked to indicate on the registration card whether or not any of their family members had previously attended the University of Buffalo. On the basis of this information it is possible to define the following three types of parental matings: (1) neither parent had previously attended the university, (2) one parent had attended the university, and (3) both parents had attended the university.

The distribution of the total 3,886 students by type of parental mating is presented in Table 4. The "gene-frequency," p, was estimated, binomial expectations were computed, and a chi-square test applied to the data. Chi-square was quite high at a probability of less than 0.001, indicating a lack of fit between expectancy and observation. On the basis of this lack of fit one would infer that attendance at the university is not determined by a recessive gene. However, since investigations of human diseases rarely deal with a group as large as 3,886 affected individuals, we considered it desirable to see what effect smaller samples would have on the probability levels. This was studied by experimentally sampling the population of university students and determining the proportion of samples that have a good fit with binomial expectations. Such a sampling experiment was carried out with random number cards and an electronic computer.

Two sample sizes were investigated: a sample size of 900 representing approximately 25 per cent of the population and one of 1,900 representing approximately 50 per cent of the population. For each sample size the computer generated 100 random samples. For each sample of each size a "p" value was computed and binomial expectations obtained. Chi-squares for each sample were determined. In Table 5 we have presented the frequency of samples by various chi-square values, together with their probability levels.

From Table 5 we note that in 54 per cent of samples of size 900 the chi-square was less than 3.841, with probability values greater than 0.05. Expressing this in terms of our genetic hypothesis, this indicates that in 54 per cent of instances a sample of 900 would have a result that was consistent with a recessive genetic hypothesis, even though the population data, from which this sample was drawn, was not consistent with a recessive gene hypothesis. When the sample size increases to 1,900 the proportion of samples that fit expectations decreases to 27 per cent at 5 per cent probability levels. In other words, given an individual sample of size 900, in which a good fit with expectation is obtained, we would be wrong about 54 per cent of the time in making a genetic inference, and with a sample size of 1,900 our chances of error would be about 27 per cent.

The lack of discriminatory ability of the chi-square test in this situation probably results from the small numbers expected in the "both parents affected" category, and this in turn is a function of the estimated gene frequency "p." Unfortunately, most genes of interest in genetic studies also have low frequencies and small expectations in the "both parents affected" category are the rule rather than the exception.

Discussion

It is evident from the foregoing that it is quite possible to obtain a good fit to at least one genetic statistical model when analyzing a characteristic that may be predominantly environmentally determined, providing that there is a degree of familial aggregation similar to that found in genetically determined conditions. The existence of such a possibility emphasizes the need for considerable caution in drawing genetic inferences from the mere existence of familial aggregation in conditions, such as cancer and psychoses, in which it has not been possible as yet to fit any definitive genetic model. The problem we have considered is a specific illustration of the difficulties encountered in determining whether a disease is inherited. This has been recently discussed by David and Snyder[3] and by Neel.[4]

The results that we have obtained are not as startling as may appear at first glance. They have been implied by Dahlberg in a recent discussion where he states that, "It should be clear from the foregoing that for statistical anal-

Table 5—Distribution of samples derived by experimental sampling by chi-square and probability values for sample sizes of 900 and 1,900

Chi-Square Values (1 d.f.)	P	Per cent Distribution of Samples	
		Sample Size	
		900	1,900
Less than 3.841	>0.05	54	27
3.842–6.635	$0.05<P<0.01$	13	11
6.636–10.827	$0.01<P<0.001$	12	14
Greater than 10.827	<0.001	21	48

yses of heredity, and particularly when the traits are at all common, we must have recourse to other and sharper methods than the mere finding that a given trait has a frequency that agrees fairly well with the classical number of 1/4."[5]

In view of these findings it is necessary to consider what additional criteria are necessary in order to increase the validity of a genetic inference. We use the term "additional" since the application of the statistical test discussed in this report may be considered a necessary first step in a genetic study. Several investigators have suggested such additional criteria. Dahlberg states that, "Far more reliable conclusions can be drawn if one studies the frequency of the trait in each category of relatives separately and then compares the actual figures with theoretical ones deduced from the formulae . . . given. . . . Then, if all the figures show good agreement random or environmental factors can hardly be responsible. . . . Such an assumption can be considered if, in addition, the frequency of the trait in the basic population is the same as the figure in column 1 of the table (in the original paper) because this figure is the basis for all the other figures. If at first we do not know the frequency of the trait-carriers in the population, we can find out by a suitable sampling procedure."[5] To these remarks must be added the statement by David and Snyder, that "the cogency of data on familial incidence or on twin concordance as evidence for the significant implication of genetic factors in the etiology of a disease rests ultimately upon the exhaustiveness with which it has been possible to exclude environmental factors as responsible for the associations found."[3]

As shown by the findings among the students of the University of Buffalo, deviations from hypothesis which are easily detected in a large sample may have only a small chance of detection in a small sample. When tests of this kind are performed it is essential that the magnitude of deviation which might reasonably go undetected be taken into account. Convincing proof that this concern is not purely academic may be found by studying other nongenetic characteristics with the standard genetic tools. Such investigations may be of further use by indicating the magnitude of deviations from genetic ratios which may generally be expected and the sample sizes needed to distinguish them from genetic ratios.

Summary

A binomial model, used for determining the consistency of observations with a recessive gene hypothesis, is shown to have a low ability to discriminate between genetic and nongenetic hypotheses. Reasons for such performance are briefly reviewed. Of particular interest is the indication that very large sample sizes of affected individuals may be necessary to reduce the chances of making erroneous inferences to an acceptable level.

REFERENCES

1. Steinberg, A. G., and Wilder, R. M. A Study of the Genetics of Diabetes mellitus. Am. J. Human Genet. 4:113 (June), 1952.
2. Allan, W. Heredity in Diabetes. Ann. Int. Med. 6:1272 (Apr.), 1933.
3. David, P. R., and Snyder, L. H. Genetics and Disease. Proc. Second Nat. Cancer Conference 2:1128, 1952.
4. Neel, J. V. On Some Pitfalls in Developing an Adequate Genetic Hypothesis. Am. J. Human Genet. 7:1 (Mar.), 1955.
5. Dahlberg, G. Genetic Investigations in Different Populations. Acta Genet. et Stat. Med. 3,2:117, 1952.

Dr. Lilienfeld is professor of public health administration and director, Division of Chronic Diseases, Johns Hopkins University School of Hygiene and Public Health, Baltimore, Md.

This paper was presented before a Joint Session of the Epidemiology and Mental Health Sections of the American Public Health Association at the Eighty-Sixth Annual Meeting in St. Louis, Mo., October 27, 1958.